BAYESIAN INFERENCE OF OPTIMAL LYMPHOSCINTIGRAPHIC SAMPLING TIMES

S. Doyle*, A. Quinn* and P. Gebouský**

* University of Dublin, Trinity College, Ireland
** ÚTIA, AVČR, Czech Republic

doylesj@tcd.ie

Abstract: Recent research has pioneered quantitative Bayesian methods for diagnosis of early-stage lymphedema. The effectiveness of this approach depends on the selection of optimal lymphoscintigraphic imaging times for the entire limb, which are currently unavailable. This paper develops a novel multi-channel parametric model of the arm and, using the Bayesian paradigm, derives an expression for the optimal lymphoscintigraphic sampling times. The Bayesian inference algorithm is applied to an over-sampled scintigraphic image sequence, and an optimal subset of such images is inferred.

Introduction

Upper-limb lymphedema is a condition involving excessive swelling of the arm. It can occur as a side-effect of breast cancer treatment and is caused by impaired circulation in the lymphatic system. Therapy is used to eliminate protein stagnation and restore normal lymphatic circulation. However, this is only effective in the early stages of the condition. If untreated, lymphatic walls undergo irreversible fibrosis and the patient may lose functionality of the limb. The ability to diagnose early stage lymphedema is therefore critical to recovery.

Lymphoscintigraphy is a non-invasive, low risk, nuclear imaging method used to diagnose early stage lymphedema [1]. Typically, 25MBq of Tc-99m-labelled human serum albumen is injected into the interstitial space of each hand. This radioactive tracer then flows through the lymph system of the arm, to the liver. Lymphoscintigraphic images record the counts from the radioactive tracer at a chosen time during its propagation through the arm. Patient, time and economic factors limit the number of imaging times. Each image is exposed for one minute, and typically only two or three images are taken. Subsequent images allow for an analysis of flow dynamics and provide an insight into the state of the lymph system.

Traditionally, these images have required qualitative examination by a medical expert. Gebouský *et al* [2] have recently developed quantitative Bayesian techniques to infer the state of the lymph system from lymphoscintigraphic images. In this Bayesian approach, the images are divided into three *Regions Of Interest* (ROIs). The radioactive counts are aggregated across each ROI, and then normalized with respect to the initial administered activity. The resulting normalized counts in each ROI correspond to the radioactive activity in each region and thus the time-dependent flow of tracer can be analysed in that region of the arm. The complicated nature of the lymphatic system, as well as abnormal lymphatic flow dynamics indicative of lymphedema [3][4], make exact modelling of the arm impossible. Gebouský *et al* model the data for each ROI independently, using a single-channel d^{th} order discrete-time delay model for each ROI. This model facilitates Bayesian inference of diagnostically significant quantities and hence a quantitative evaluation of the underlying lymphatic condition.

The approach is sensitive to the temporal location of imaging times. The single-channel model is localized to individual ROI, and therefore is incapable of determining the optimal sampling times across all three ROI. In this paper, we present a multi-channel parametric model of the arm which models the normalized counts from all three ROI simultaneously. This model is then used to infer the optimal sampling times across all ROIs.

Multi-Channel Parametric Model

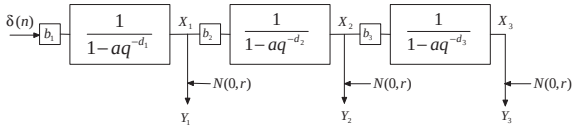

Figure 1: Multi-Channel Parametric Model of Normalized ROI Counts. q^{-1} is the backward shift operator ($q^{-1}x_t = x_{t-1}$)

The counts for each ROI are modeled with a discrete-time d^{th} order delay model. The d^{th} order delay model is also employed in the single-channel model. It can realize a rich ensemble of possible curves, capturing the stable, slowly-decaying, non-oscillatory nature of the observed lymphoscintigraphic counts. The delay models are cascaded to reflect the gradual propagation of radioactive tracer through the lymph system of the arm (Figure 1). Here, the forearm, upper-arm and axilla ROI yield three channels of data, denoted by $c = 1, 2, 3$ respectively.

The model is parameterized by $\theta = (b_1, b_2, b_3, \vartheta, r)$,

ISSN: 1727-1983

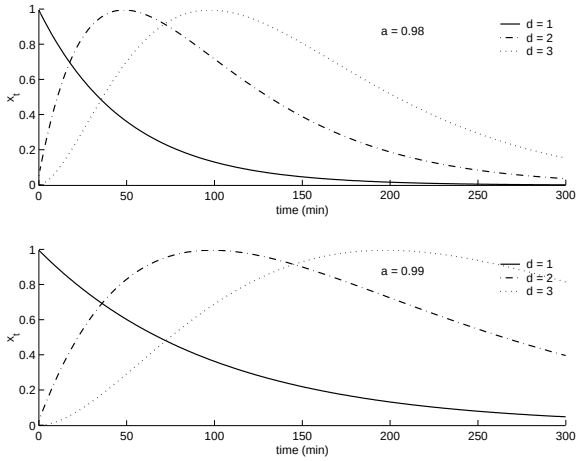

Figure 2: Sample Time Activity Curves for the d^{th} Order Delay Model (*normalized so that Curve Maximum = 1*)

where $\vartheta = (a, d_1, d_2, d_3)$. Gain parameters, b_1, b_2 and b_3, are used to adjust signal amplitude and allow for an expanded set of possible system responses. The partial time constant, a, is such that the system response is stable and non-oscillatory. A sufficiently large set of candidate response curves allow for a to be assumed identical for each channel. The model orders, d_1, d_2 and d_3, describe the propagation rate of the tracer through the lymph system of the arm.

The injection of radioactive tracer is modeled by a *Unit Sample Sequence* (USS) $u[t]$ at the system input. The discrete-time output at each ROI is sampled at time $t \in \mathcal{T} = \{1, 2, \ldots, N\}$, where N is the cardinality of the complete set of sampling times, $\mathcal{T}$. The USS response of channel c is known as the *Time Activity Curve* (TAC), $X_c^{\mathcal{T}} = \{x_{c,1}, x_{c,2}, \ldots, x_{c,N}\}^T$. Typical TAC are shown in Figure 2. Superscript $\mathcal{T}$ is used to denote the vector of times at which a ROI is observed. ^{T} universally denotes transposition. Causality requires that $x_{c,t} = 0, \forall t < 0$. The closed form solution [5] of the USS response of channel c is given by

$$x_{c,t} = b_c \binom{t + d_c - 1}{t} a^t u[t], \qquad t \in \mathbb{Z}^+. \quad (1)$$

The gain parameter b_c enters the analysis linearly, while the non-linear terms are $\vartheta_c = (a, d_c)^T$. Thus $X_{c,t} = b_c A_{\vartheta_c,t}$ where

$$A_{\vartheta_c,t} = \binom{t + d_c - 1}{t} a^t u[t], \qquad t \in \mathbb{Z}^+. \quad (2)$$

The cascade is tapped after each ROI, and noise is superimposed on the signal by the observation process. The data $\mathbf{Y} = \{Y_1, Y_2, Y_3\}$ are the observed normalized counts from channel 1, 2 and 3 respectively, with added noise.

Independence of samples can be assumed under sparse sampling conditions. Aggregation within each ROI allows the Poisson distributed noise on individual counts to be approximated well by zero-mean Gaussian noise $e_{c,t}$. $e_{c,t}$ are i.i.d. $\forall t \in \mathcal{T}$ owing to the large sampling period of $\mathcal{T}$. Furthermore, the channel noise processes may be assumed to be mutually independent. Thus the probability density function (pdf) of the observation vector $Y_c^{\mathcal{T}} = \{y_{c,1}, y_{c,2}, \ldots, y_{c,N}\}^T$, where $y_{c,t} = x_{c,t} + e_{c,t}$, is given by

$$p(Y_c^{\mathcal{T}}|\theta,\mathcal{T}) = (2\pi r)^{-0.5N} \exp\left[\frac{-1}{2r}||Y_c^{\mathcal{T}} - b_c A_{\vartheta_c}^{\mathcal{T}}||^2\right], \quad (3)$$

$||\cdot||$ denotes the Euclidean norm. The variance of the noise, r, is assumed constant $\forall c, t$. The pdf of the observation vector for the entire limb, $\mathbf{Y}$, is derived from Equation (3).

$$p(\mathbf{Y}^{\mathcal{T}}|\theta,\mathcal{T}) = l(\theta|\mathcal{T},\mathbf{Y}^{\mathcal{T}}) = (2\pi r)^{-1.5N} \exp\left[\frac{-1}{2r}\sum_{c=1}^{3}||Y_c^{\mathcal{T}} - b_c A_{\vartheta_c}^{\mathcal{T}}||^2\right], \quad (4)$$

Where $l(\theta|\mathcal{T},\mathbf{Y}^{\mathcal{T}})$ denotes the Likelihood Function (LF) of θ.

Prior Information

The Bayesian approach demands that the probability distributions for model parameters must be determined. Gebouský [6] elicits prior distributions for the parameters of the single channel model, and these are used to determine the prior distributions for multi-channel model parameters.

a is dependant on d_c, and is typically observed in the range $0.9 \leq a \leq 0.999$. d_c describes the rate of propagation of the radioactive tracer and is observed in the range $1 \leq d_c \leq 5$. The multi-channel model indicates that $d_{c+1} = d_c + d, \quad d \in \mathbb{Z}^+$, with $d_1 \in \{1, 2, 3\}$, $d_2 \in \{2, 3, 4\}$ and $d_3 \in \{3, 4, 5\}$. a and d_c determine the location of the TAC maximum, which is observed in the range $(25, 250)$ for channel 1, $(60, 300)$ for channel 2 and $(100, 360)$ for channel 3.

b_c determines the amplitude of the USS response, and depends on the non-linear parameters ϑ_c. The data suggest that the maximum of the response does not exceed 20% of the applied input, and that it is non-negative. $b_c \in (0, \bar{b}_{\vartheta_c})$ where $\bar{b}_{\vartheta_c}$ is the maximum value of b_c and is evaluated numerically for each ϑ_c. The noise is an artefact of the measurement process and as such is patient and limb independent. The variance of the noise is also independent of all other parameters and empirically is in the range $r \in (10^{-6}, 10^{-4})$. The prior distribution $p(\theta)$ can thus be expressed as a product of conditional distributions given the aforementioned dependencies.

$$p(\theta) = p(b_1|\vartheta, r)p(b_2|\vartheta, r)p(b_3|\vartheta, r)p(\vartheta)p(r) \quad (5)$$

ISSN: 1727-1983

The parameter a is discretized to facilitate computation. A uniform pdf on $\vartheta = (a, d_1, d_2, d_3)^T$ is invoked according to the *principle of insufficient reason* [7]. The pdfs on the continuous parameters b_c and r are conjugate to Equation (4) [8].

$$p(b_c|\vartheta, r) = (2\pi\omega_{c,\vartheta} r)^{-0.5} \exp\left[\frac{-1}{2\omega_{c,\vartheta} r}||b_c - \hat{b}_{c,\vartheta}||^2\right] \quad (6)$$

$\omega_{c,\vartheta} > 0$ and $\hat{b}_{c,\vartheta}$ parameterize this distribution. The expected range of b_c is used to determine the relationships $\hat{b}_{c,\vartheta} = 0.5\bar{b}_{c,\vartheta}$ and $\omega_{c,\vartheta} = \bar{b}^2_{c,\vartheta}/(4\hat{r})$, where $\hat{r}$ is a conservative estimate of the measurement variance r, $\hat{r} = 10^{-5}$.

$$p(r|\delta, \varepsilon) = (\kappa)^{0.5\delta}\,\Gamma^{-1}\,(0.5\delta)\,r^{-0.5\delta-1}\exp\left[-\frac{\kappa}{r}\right] \quad (7)$$

$$\text{where,} \qquad \kappa = \frac{(\delta-2)\varepsilon}{2}$$

δ and ε are known hyperparameters. $\Gamma(\cdot)$ is the Euler gamma function. In [2], a Gaussian approximation of (7) is used with the one-standard-deviation confidence interval set by physical considerations. This leads to the choice $\delta = 7,\ \varepsilon = 3 \times 10^{-5}$. From Equations (5), (6), and (7):

$$p(\theta) \propto p(\vartheta) r^{-\frac{1}{2}(\delta+5)} \prod_{c=1}^{3}\left(\omega_{c,\vartheta}^{-0.5}\right) \times \exp\left[\frac{-1}{2r}\left((\delta-2)\varepsilon + \frac{1}{\omega_{c,\vartheta}}\sum_{c=1}^{3}||b_c - \hat{b}_{c,\vartheta}||^2\right)\right] \quad (8)$$

The posterior inference of the model is found by multiplying Equations (4) and (8).

$$p(\theta|\mathbf{Y}^{\mathcal{T}}, \mathcal{T}) \propto p(\mathbf{Y}^{\mathcal{T}}|\theta, \mathcal{T})p(\theta) \quad (9)$$

Optimal Sampling Grid Selection

We consider the set of possible times, $\mathcal{T} = \{1, 2, \ldots, N\}$ to be the common superset of every possible sparse sampling grid $\mathcal{S} = \{t_i, \ldots, t_{N_\mathcal{S}}\}$; i.e. $t_i \in \mathcal{T},\ N_\mathcal{S} \leq N$. Typically $N_\mathcal{S} << N$, eg. 2 or 3. The number of possible choices of $\mathcal{S}$ is therefore ${}^N C_{N_\mathcal{S}}$.

Since only a small number of images may be recorded, the choice of the recording times is critical. Currently, the sparse set of sampling times $\mathcal{S} \subseteq \mathcal{T}$ are chosen empirically. Our task now is to infer $\mathcal{S}$ via the multi-channel parametric model. This should provide an optimal inference for the entire limb, rather than a ROI-specific inference.

To ensure the reportability of the inference, it is necessary that there exist a database of patient data that sufficiently represents the ensemble of all possible patient-specific realizations. $\mathcal{Y}^{\mathcal{T}} = \{\mathbf{Y}_1^{\mathcal{T}}, \ldots, \mathbf{Y}_P^{\mathcal{T}}\}$ is the set of all patient data, where P is the number of patients.

We wish to select the subset of sampling times $\mathcal{S}$ for which $p(\mathcal{S}|\mathcal{Y}^{\mathcal{T}})$ is maximised. Bayes' theorem yields $p(\mathcal{S}|\mathcal{Y}^{\mathcal{T}}) \propto p(\mathcal{Y}^{\mathcal{T}}|\mathcal{S})p(\mathcal{S})$. A uniform distribution is invoked on $p(\mathcal{S})$, and the *Maximum A Posteriori* (MAP) value of $\mathcal{S}$, $\mathcal{S}_{\text{MAP}}$, is given by

$$\mathcal{S}_{\text{MAP}} = \arg\max_{\mathcal{S}} p(\mathcal{Y}^{\mathcal{T}}|\mathcal{S})$$

Patient-specific characteristics are independent so

$$\mathcal{S}_{\text{MAP}} = \arg\max_{\mathcal{S}} \prod_{i=1}^{P} p(\mathbf{Y}_i^{\mathcal{T}}|\mathcal{S})$$

$p(\mathbf{Y}_i^{\mathcal{T}}|\mathcal{S})$ is the patient-specific predictive pdf conditioned on the sub-sampled times $\mathcal{S}$.

$$p(\mathbf{Y}_i^{\mathcal{T}}|\mathcal{S}) = \int_\theta p(\mathbf{Y}_i^{\mathcal{T}}|\theta, \mathcal{T})p(\theta|\mathbf{Y}_i^{\mathcal{S}}, \mathcal{S})d\theta$$

Where $p(\mathbf{Y}_i^{\mathcal{T}}|\theta, \mathcal{T})$ is the likelihood function given by Equation (4). $p(\theta|\mathbf{Y}_i^{\mathcal{S}}, \mathcal{S})$ is the posterior distribution evaluated on a candidate sampling grid $\mathcal{S}$. From Bayes' theorem:

$$p(\theta|\mathbf{Y}_i^{\mathcal{S}}, \mathcal{S}) \propto p(\mathbf{Y}_i^{\mathcal{S}}|\theta, \mathcal{S})p(\theta)$$

Here, $p(\mathbf{Y}_i^{\mathcal{S}}|\theta, \mathcal{S})$ is the likelihood function (4) evaluated on the vector of times $\mathcal{S}$. Equation (8) gives the expression for $p(\theta)$. Thus $p(\mathbf{Y}_i^{\mathcal{T}}|\mathcal{S})$ can be elicited and $\mathcal{S}_{\text{MAP}}$ reveals the optimal sampling times.

Sampling Grid Inference Experiments

Recall that the ultimate aim is to infer reliably the parameters of an unseen patient case. Hence, an effective sub-selection $\mathcal{S}$ will be one for which the associated data $\mathcal{Y}^{\mathcal{S}}$ is most sensitive to changes in θ.

Observation of the data reveals that few counts are recorded in the initial period after injection, and the curves for different parameters are difficult to distinguish. A period of high activity then prevails as the radio-active tracer propagates through the ROI. The activity then diminishes and curves for different parameters once again coincide. The TAC are most sensitive to changes in θ in the high-activity region, and the optimal sampling times should reflect this fact.

Simulated Data Case

Artificial lymphoscintigraphic data is generated using the multi-channel parametric model. Realizations of θ from Equation (8) are used to generate data for

ISSN: 1727-1983

Table 1: Sampling Combinations and associated Probability for Simulated Data. (*Sampling Combinations for which* $p(\mathcal{S}|\mathcal{Y}^{\mathcal{T}}) < 10^{-4}$ *are not displayed*)

S	$p(\mathcal{S}\|\mathcal{Y}^{\mathcal{T}})$
$\{70, 190\}$	0.907
$\{70, 210\}$	0.078
$\{90, 210\}$	0.01
$\{90, 190\}$	0.003

three-thousand *virtual* patients. Notionally then, the database is representative of the ensemble of possible simulated patient records. The complete sampling set is $\mathcal{T} = \{10, 30, 50, 70, 90, 110, 130, 150, 170, 190, 210\}$. The number of sampling times is $N = 11$, while the number of sub-sampled times is $N_{\mathcal{S}} = 2$, allowing for a comprehensive choice of ${}^{11}C_2 = 55$ possible combinations. $p(\mathcal{S}|\mathcal{Y}^{\mathcal{T}})$ is evaluated for each choice of $\mathcal{S}$ and the results are shown in Table 1. Of the fifty-five possible sampling time combinations, $\mathcal{S}_{\text{MAP}} = \{70, 190\}$ is maximum. Note from Figure 2 that the optimal sampling times for simulated data are in the high-activity region of many of the candidate curves.

To verify the ability of the inference procedure to select the optimal sampling times, we observe the location of the sampling times as the high-activity regions in the data vary according to the realization of θ. The maximum of the TAC for each channel is selected as an indicator of the location of the high-activity region.

The average of the optimal sampling times is plotted against the mean time to maximum of the three channels. The result is shown in Figure 3. As the mean time to maximum of channel 1, 2 and 3 increases, the average of the optimal sampling times increases. The correlation coefficient is 0.926. Thus the optimal sampling times track the high-activity characteristic region of the data.

Real Data Case

The analysis is fed with real data gathered from the lymphoscintigraphic imaging of sixteen patients, or thirty-two limbs. Each patient has undergone breast-cancer treatment, and has suspected early-stage lymphedema in one or more limbs. The number of sampling times for each patient is typically three or four. The actual sampling times used differ slightly, and so data is grouped into four *sampling intervals* to facilitate the analysis. The intervals are denoted by $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$ and $\mathcal{I}_4$. The selection of optimal sampling times becomes a selection of optimal sampling intervals. A choice of ${}^4C_2 = 6$ possible combinations exist for $\mathcal{S}$. The optimal sampling intervals for real data are shown in Table 2. Of the six possible sampling interval combinations, the selection $\{\mathcal{I}_2, \mathcal{I}_4\}$ is maximum a posteriori. The optimal sampling intervals for real data are $\langle 65, 105\rangle$ and $\langle 170, 220\rangle$.

Recall that the optimal sampling times for simulated data are 70 minutes and 190 minutes. These times fall within the optimal sampling intervals for real data of

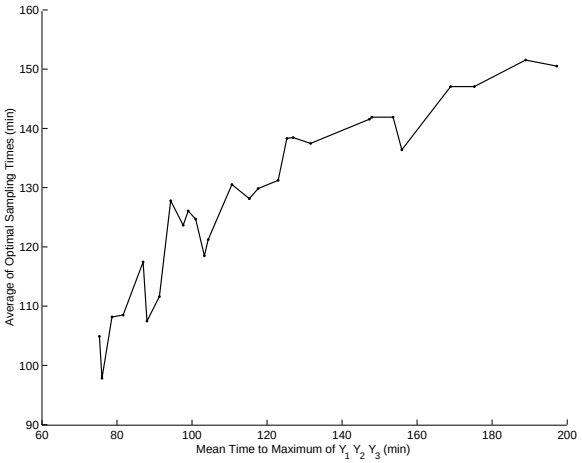

Figure 3: Average of $\mathcal{S}$ against Mean Time to Maximum of $\mathcal{Y}$

Table 2: Optimal Sampling Intervals for Real Data

Interval	Range (mins)	$\mathcal{S}_{\text{MAP}}$
I_1	$\langle 20, 60\rangle$	
I_2	$\langle 65, 105\rangle$	•
I_3	$\langle 110, 160\rangle$	
I_4	$\langle 170, 220\rangle$	•

$\langle 65, 105\rangle$ and $\langle 170, 220\rangle$. This result implies the validity of the multi-channel model, the prior information and the simulated data.

Note that the MAP selection of $\{\mathcal{I}_2, \mathcal{I}_4\}$ for real data places the sampling intervals near the high-activity region of many of the candidate curves. We note from Figure 2 that this region is also one of high variability. The candidate curves at the location of the optimal sampling intervals are highly sensitive to changes in θ. This fulfills the aim of optimal sampling time inference, given at the beginning of this section.

Conclusion

The selection of optimal sampling times is critical for accurate diagnosis of early-stage lymphedema. This paper presents a Bayesian technique for inference of the optimal sampling times for the entire limb.

The multi-channel parametric model describes the time-varying lymphoscintigraphic counts from all ROIs simultaneously. The optimal sampling times are therefore a compromise for all ROIs. An expression for the predictive distribution $p(\mathcal{Y}^{\mathcal{T}}|\mathcal{S})$ is derived. Real lymphoscintigraphic data is used to evaluate the optimal sampling intervals of $\langle 65, 105\rangle$ and $\langle 170, 220\rangle$ for a patient database. Artificial data generated on the expected range of θ yields the optimal sampling times of 70 and 190 minutes. This validates the multi-channel parametric model, and the Bayesian inference of $\mathcal{S}_{\text{MAP}}$.

The optimal sampling times allow for a more accurate diagnosis of early-stage lymphedema using quanti-

ISSN: 1727-1983

tative methods, facilitating early therapeutic treatment of the limb and thereby helping to preserve its functionality.

Acknowledgements

This work is funded by the Irish Research Council for Science, Engineering and Technology (IRCSET)

References

[1] P. BOURGEOIS. Imaging techniques in the management and prevention of posttherapeutic upper limb edemas. *Cancer*, 83(12):2805–2813, 1998.

[2] PETR GEBOUSKÝ, MIROSLAV KÁRNÝ, and ANTHONY QUINN. Lymphoscintigraphy of Upper Limbs: A Bayesian Framework. *Selected Proceedings of the 7th Valencia International Meeting on Bayesian Statistics*, 2003.

[3] A.W. STANTON, W.E. SVENSSON, R.H. MELLOR, A.M. PETERS, J.R. LEVICK, and P.S. MORTIMER. Differences in Lymph Drainage between Swollen and Non-Swollen Regions in Arms with Breast-Cancer-Related Lymphedema. *Clin Sci (Lond)*, 101(2):131–140, Aug 2001.

[4] H. VERLOOY, JP. BISCOMPTE, I. NIEUBORG, P. DRENT, C. SCHIEPERS, L. MORTELMANS, and M. DE ROO. Noninvasive evaluation of lymphovenous anastomosis in upper limb lymhedema. *The European Journal of Lymphology*, 6(21):27–33, 1997.

[5] ALAN V. OPPENHEIM, RONALD W. SCHAFER, and JOHN R. BUCK. *Discrete-time signal processing (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.

[6] PETR GEBOUSKÝ. *Bayesian Quantitative Lymphoscintigraphy of Upper Limbs*. Ph.D. dissertation, University of West Bohemia, Pilsen, 2004.

[7] S. JAMES PRESS. *Subjective and objective Bayesian statistics. Principles, models, and applications (2nd ed.)*. Wiley Series in Probability and Statistics, 2003.

[8] J.M.BERNARDO and A.F.M.SMITH. *Bayesian Theory*. Wiley, 1994.

ISSN: 1727-1983